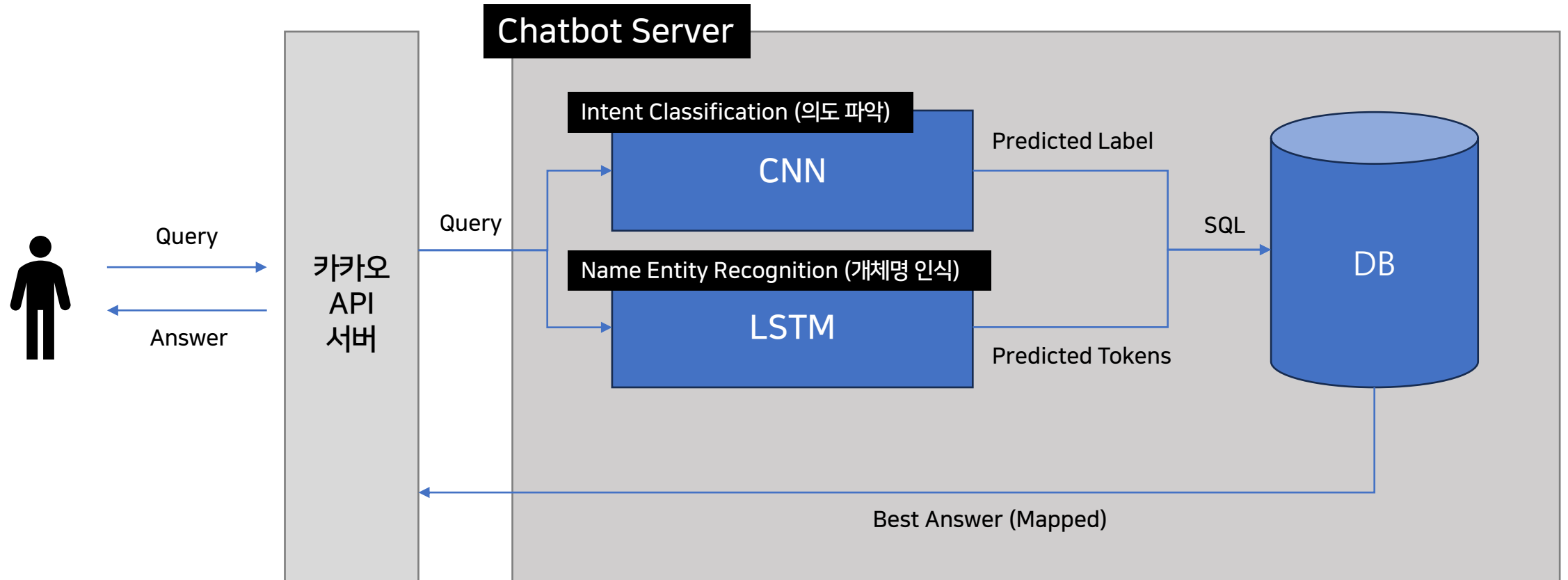


CNN, LSTM 기반 철도 질의응답 챗봇

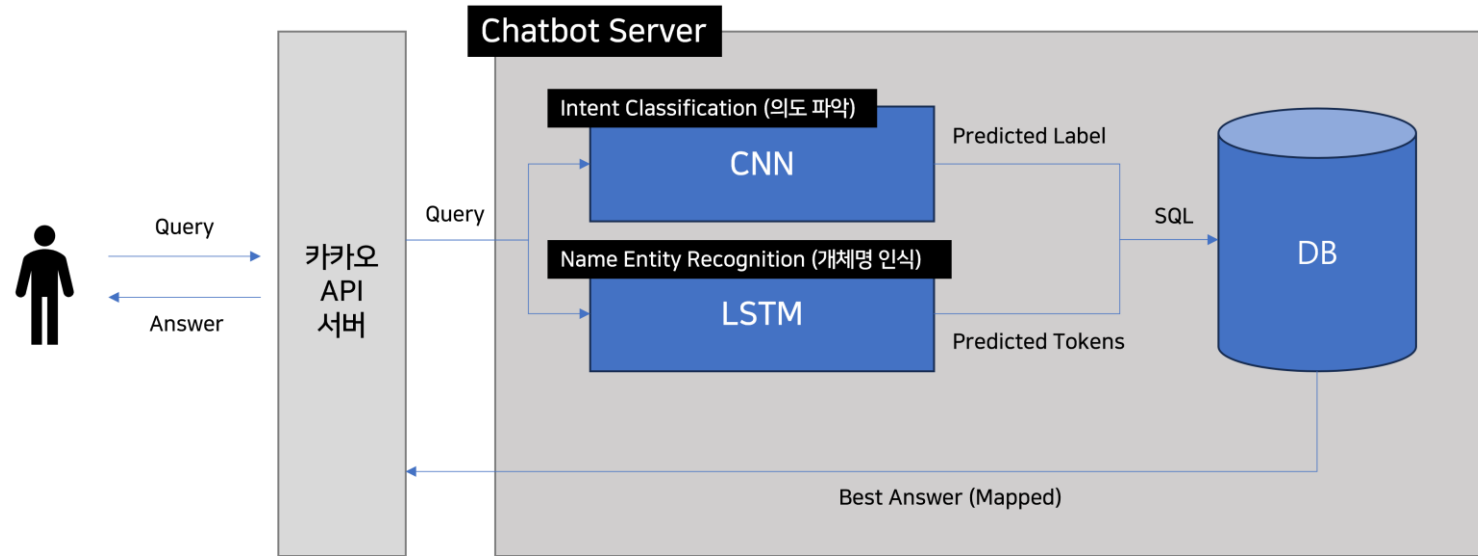
목차

1. 프로젝트 구조
2. CNN, LSTM
3. 개발 진행상황
4. 한계점 및 개선사항

1. 프로젝트 구조



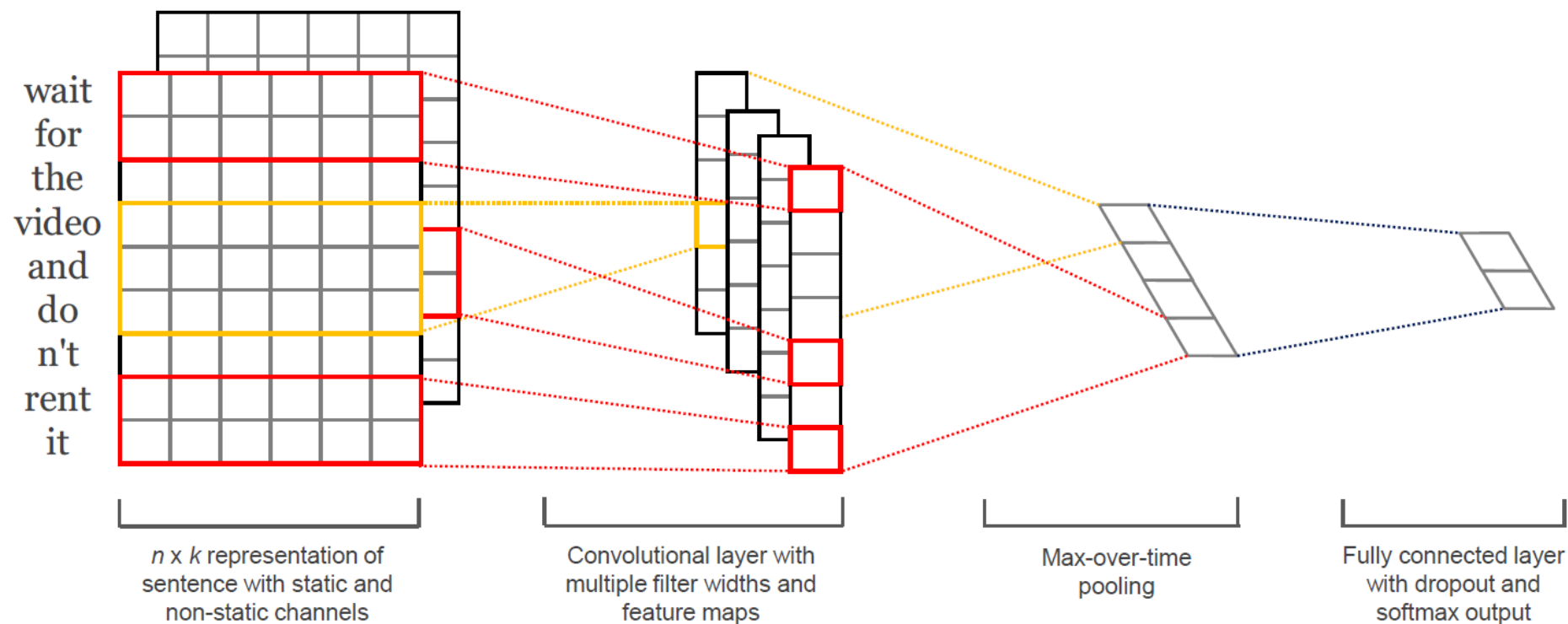
1. 프로젝트 구조



1. 카카오 API를 바탕으로 Client의 Input(질문;Query)를 받는다.
2. Query를 CNN, LSTM 모델에 넣어 각각의 결과값을 얻는다.
3. SQL문으로 결과값 조건에 알맞은 답변을 도출한다.
4. 카카오 API 서버에 결과값을 넘겨주고 Client에게 답변을 전달한다.

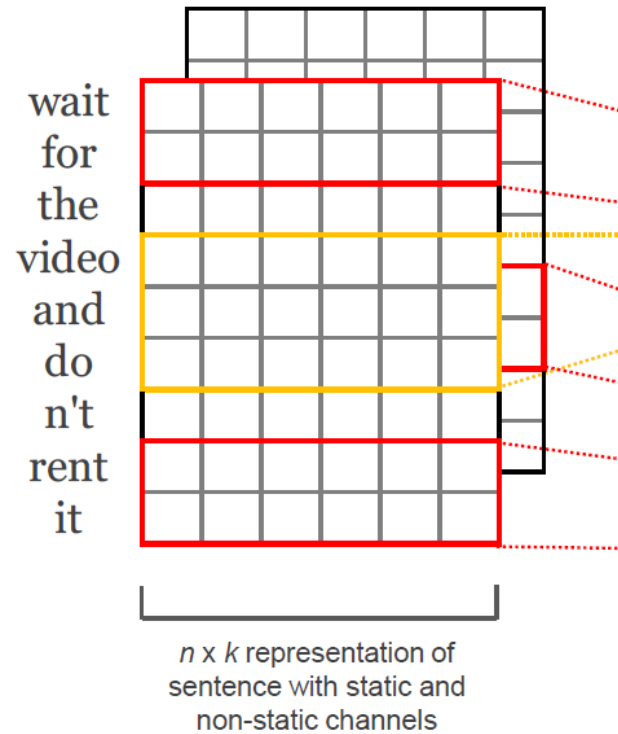
2. CNN, LSTM

CNN for Sentence Classification



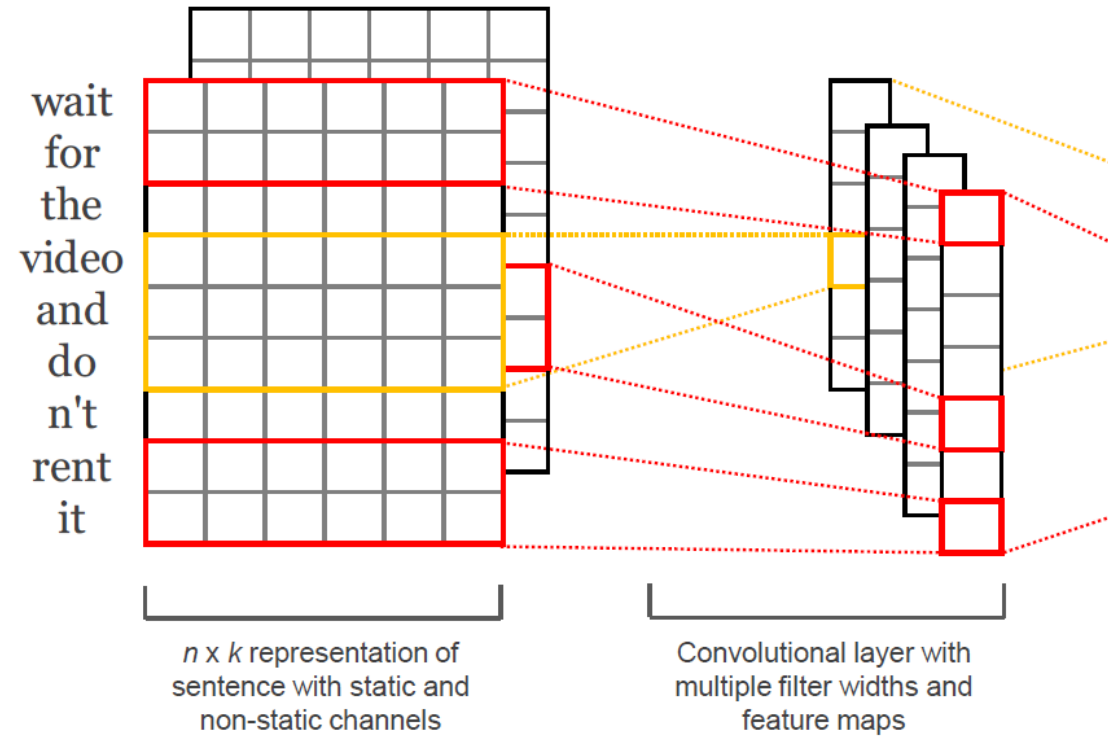
[Yoon Kim\(2014\)](#), Convolutional Neural Networks for Sentence Classification

2. CNN, LSTM



n 개의 단어로 이루어진 문장 t 개(t 차원)를 Embedding하여 벡터로 변환
(이때, 문장이 n 개의 단어로 이루어지지 않을 수 있는데 이는 Padding으로 처리)

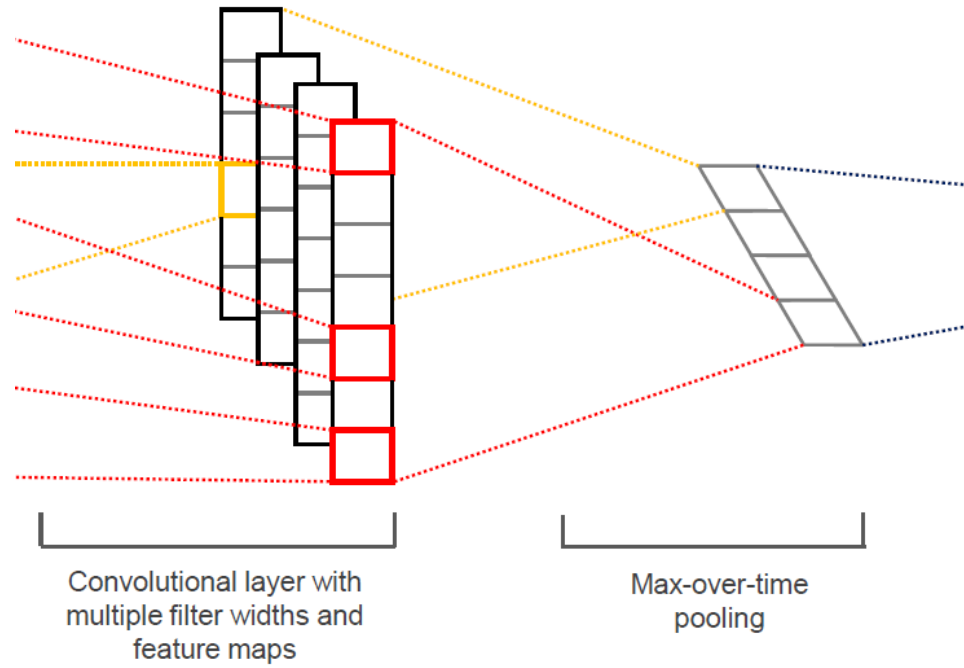
2. CNN, LSTM



n개의 단어 벡터를 Filter에 통과시켜 concatenation 해준다

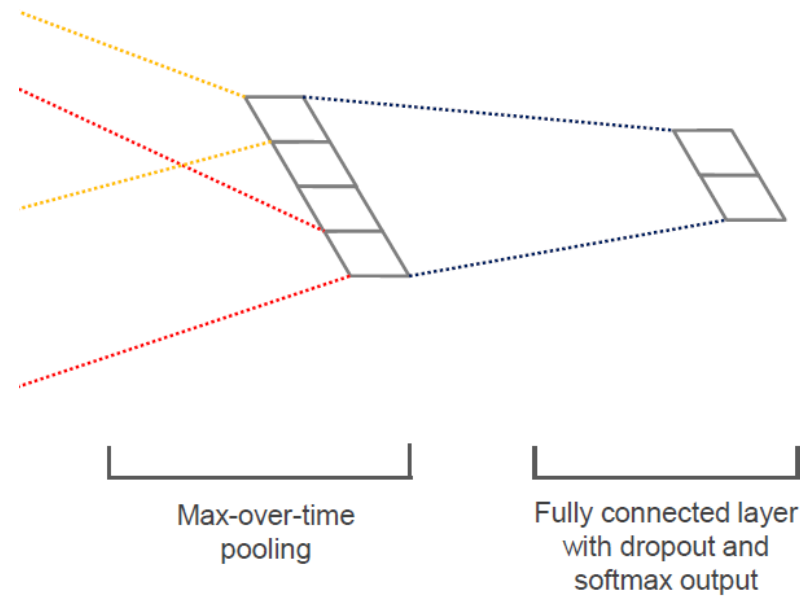
그러면 길다란 1차원 벡터 (Feature Map)을 얻을 수 있다

2. CNN, LSTM



Feature Map 중에 가장 큰 값만을 남긴다 (Max-over-time Pooling)

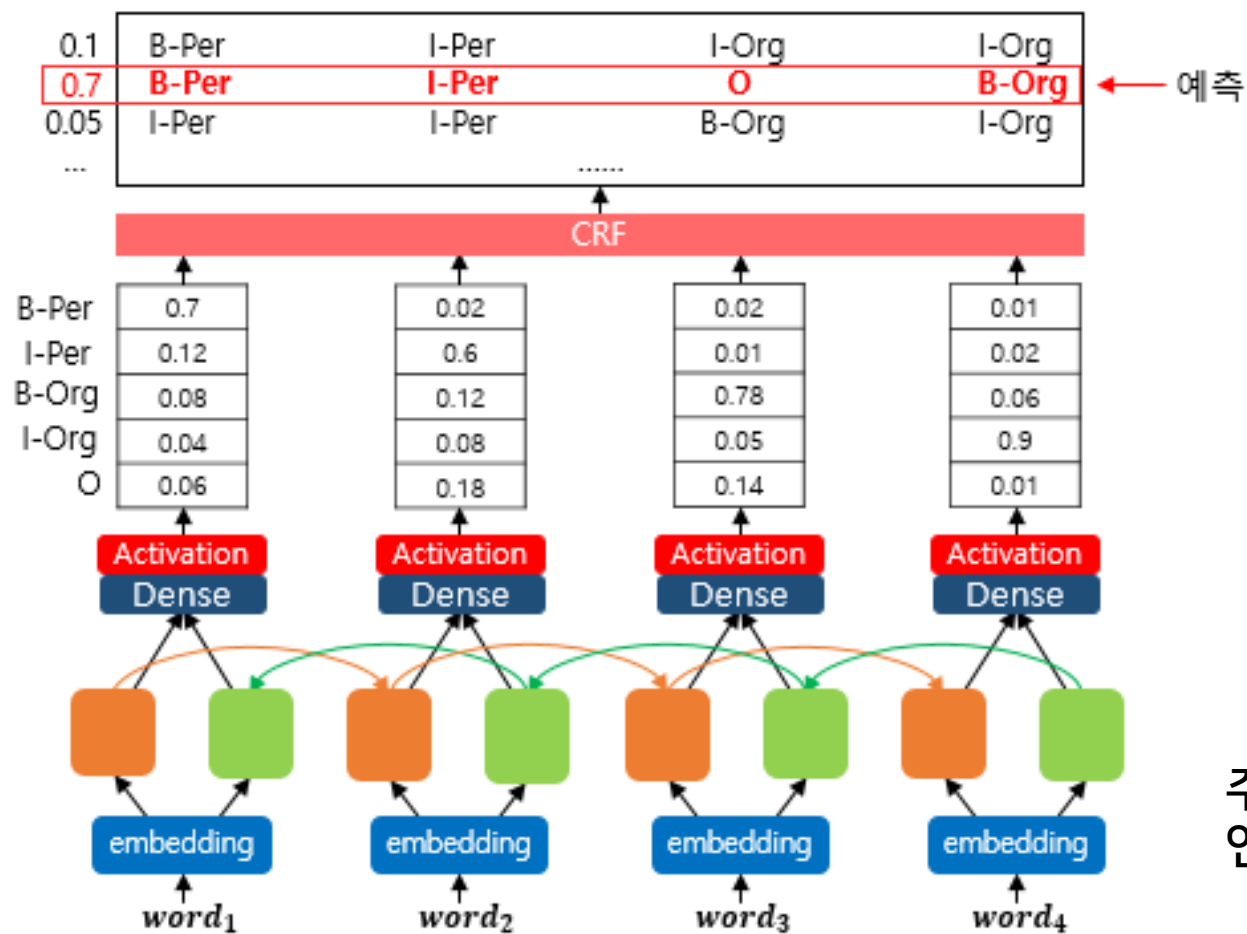
2. CNN, LSTM



Full connected softmax layer 통과 (Output)

2. CNN, LSTM

LSTM-CRF for Named Entity Recognition



3. 개발 진행상황

- 1) 데이터 확보 및 전처리를 통한 데이터셋 구축 (진행중)
- 2) CNN, LSTM 모델 개발 (완료)
- 3) Database 구축 (7월 둘째 주, 진행중)
- 4) Server 개발, Client - Server 테스트 (7월 셋째 주, 진행예정)
- 5) Server와 카카오 API 연동 (7월 넷째 주, 진행예정)
- 6) 테스트 서비스 배포 및 피드백 (7월 말 ~ 8월 초, 진행예정)

3. 개발 진행상황



QA Chatbot.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 7월 5일에 마지막으로 수정됨

+ 코드 + 텍스트

QA Chatbot

Structure

{개발중: 완료 X, 디벨롭 필요: 완성O[보수 필요], 주석없음: 완성 O}

```
chatbot
├── config
├── train_tools
│   └── dict
├── utils
├── model
│   ├── intent
│   └── ner
├── test
└── sample
```

```
[ ] !pip install tensorflow
    !pip install keras
    !pip install konlpy
    !pip install seqeval
```

접근성과 효율성을 높인 모델 학습을 위한 Colab으로의 소스코드 이전

3. 개발 진행상황

ASCL_chatbot\ascl_chatbot\chatbot_train_data\ - HeidiSQL 12.3.0.6589

파일 편집 검색 쿼리 도구 이동 도움말

데이터베이스 테이블 필터

호스트: 127.0.0.1 데이터베이스: ascl_chatbot 테이블: chatbot_train_data 데이터 쿼리

ascl_chatbot.chatbot_train_data: 28 행 (중) (대략적)

id	intent	NER	answer
1	공통	B_OG	철도자랑 · 용품 활성화인 및 개조수업검사 시 일용자...
2	공통	B_FOOD	입회는 신청자가 요청하는 시험의 유효성을 확인하기 ...
3	공통	B_DT	신청자 요청에 긴급하게 대응할 수 있도록 필요 시, 전...
4	공통	B_FOOD	검사기관에서는 이의제기 및 불만처리 절차를 시행하...
5	공통	B_OG	기술기준 관련 기술책임자 이력이 발생한 경우, 해당 ...
6	공통	I	배정받은 기술책임자가 마무리하는 것이 원칙이지만, ...
7	공통	B_PS	활성수업에는 국가가 고시한 기술기준을 적용하여야 ...
8	공통	B_LC	통발력인서 발급의 경우, 이미 관련 개별 확인서의 입...
9	공통	NNP	상호 인정이 가능한 국가 인증기관에서 발급된 것, 공...
10	공통	B_TI	발주계약일 기준의 최신 고시일로 적용합니다.
11	공통	PAD	검사 수수료는 대상(철도자랑·철도용품·철도개조)마...
12	공통	O	유효기간 이내의 공기관 시험성적서는 사전기술검토...
13	철도자랑	I	이미 활성화인을 받은 철도자랑과 동일한 자량의 경우...
14	철도용품	B_DT	하위품목들도 활성화인을 받은 용품을 사용하여 시스...
15	철도용품	B_FOOD	결과물이 다른 활성화 바뀌다고 볼 수 있습니다. 자...
16	철도용품	B_OG	활성수업 받은 용품은 자량 활성화인 시 활성화인 용...
17	철도용품	B_OG	활성수업 및 제작자수인 모두 받아야 합니다. 비용은 ...
18	철도용품	B_PS	활성변경 신고 등의 제도를 이용하여 수수료를 감면할...
19	철도용품	B_LC	설계특이성으로 인해 기술기준을 적용하기 어려운 경...
20	철도용품	NNP	철도안전법 제26조제4항제4호에 따른 특수자에 사용...
21	철도용품	R TT	철도자랑의 활성화인제도 시행 이전부터 운영중인 활...

필터: 정규 표현식

```
22 SELECT * FROM `information_schema`.`COLUMNS` WHERE TABLE_SCHEMA='ascl_chatbot' AND TABLE_NAME='chatbot_train_data' ORDER BY ORDINAL_POSITION;
23 SHOW INDEXES FROM `chatbot_train_data` FROM `ascl_chatbot`;
24 SELECT * FROM information_schema.REFERENTIAL_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='ascl_chatbot' AND TABLE_NAME='chatbot_train_data' AND REFERENCED_TABLE_NAME='chatbot_train_data';
25 SELECT * FROM information_schema.KEY_COLUMN_USAGE WHERE TABLE_SCHEMA='ascl_chatbot' AND TABLE_NAME='chatbot_train_data' AND REFERENCED_TABLE_NAME IS NOT NULL;
26 SHOW ENGINES;
27 SHOW COLLATION;
28 SHOW CREATE TABLE `ascl_chatbot`.`chatbot_train_data`;
29 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='ascl_chatbot' AND TABLE_NAME='chatbot_train_data';
30 SELECT * FROM `ascl_chatbot`.`chatbot_train_data` LIMIT 1000;
```

r1: c1 연결된: 00:00 h MariaDB 11.1.0 가동 시간: 7 일 08:36 h 서버 시간: 오후 유휴

테스트용 Database 구축 (Maria DB) + SQL문 작성 완료

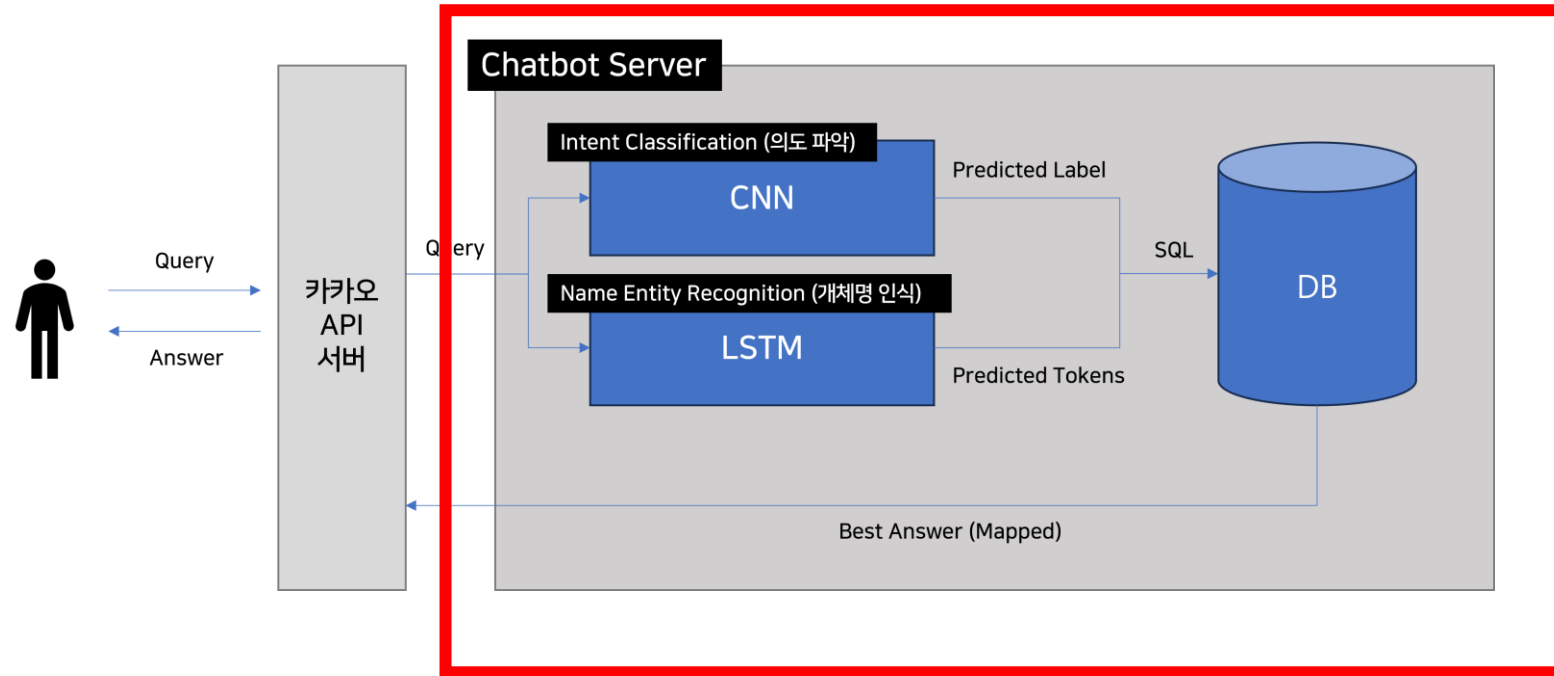
3. 개발 진행상황

```
chatbot_test x
C:\Users\slaye\anaconda3\python.exe C:\Users\slaye\OneDrive\Desktop\챗봇\chatbot\test\chatbot_test.py
질문을 입력해 주세요: 시설 유지보수 및 사고복구 등 특수목적의 철도차량에 사용되는 철도용품도 형식승인 대상인가요?
1/1 [=====] - 0s 108ms/step
1/1 [=====] - 0s 341ms/step
1/1 [=====] - 0s 19ms/step
질문 : 시설 유지보수 및 사고복구 등 특수목적의 철도차량에 사용되는 철도용품도 형식승인 대상인가요?
=====
의도 파악 : 철도용품
개체명 파악 : [('시설', 'I'), ('유지', 'O'), ('보수', 'B_PS'), ('및', 'O'), ('사고', 'O'), ('복구', 'O'), ('등', 'O'), ('특수', 'O'), ('목적', 'O'), ('철도차량', 'O'), ('사용', 'O')]
답변 검색에 필요한 NER 태그 : ['I']
=====
답변 : 철도차량의 형식승인제도 시행 이전부터 운영중인 철도차량 및 철도 용품의 유지 및 보수용으로만 사용되는 용품 중에서 다른 철도차량 및 철도용품과 호환되지 않는 용품은 형식승인 대상에서 제외됩니다.
저장 완료

Process finished with exit code 0
```

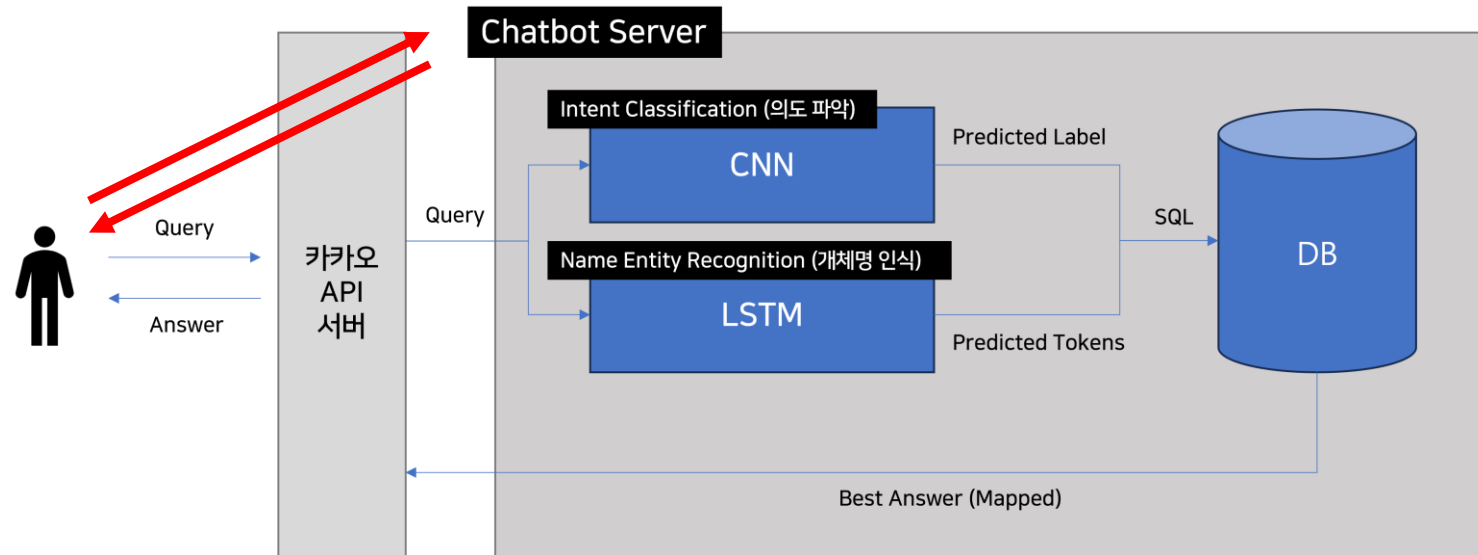
모델 Test

3-1. 개발 진행예정



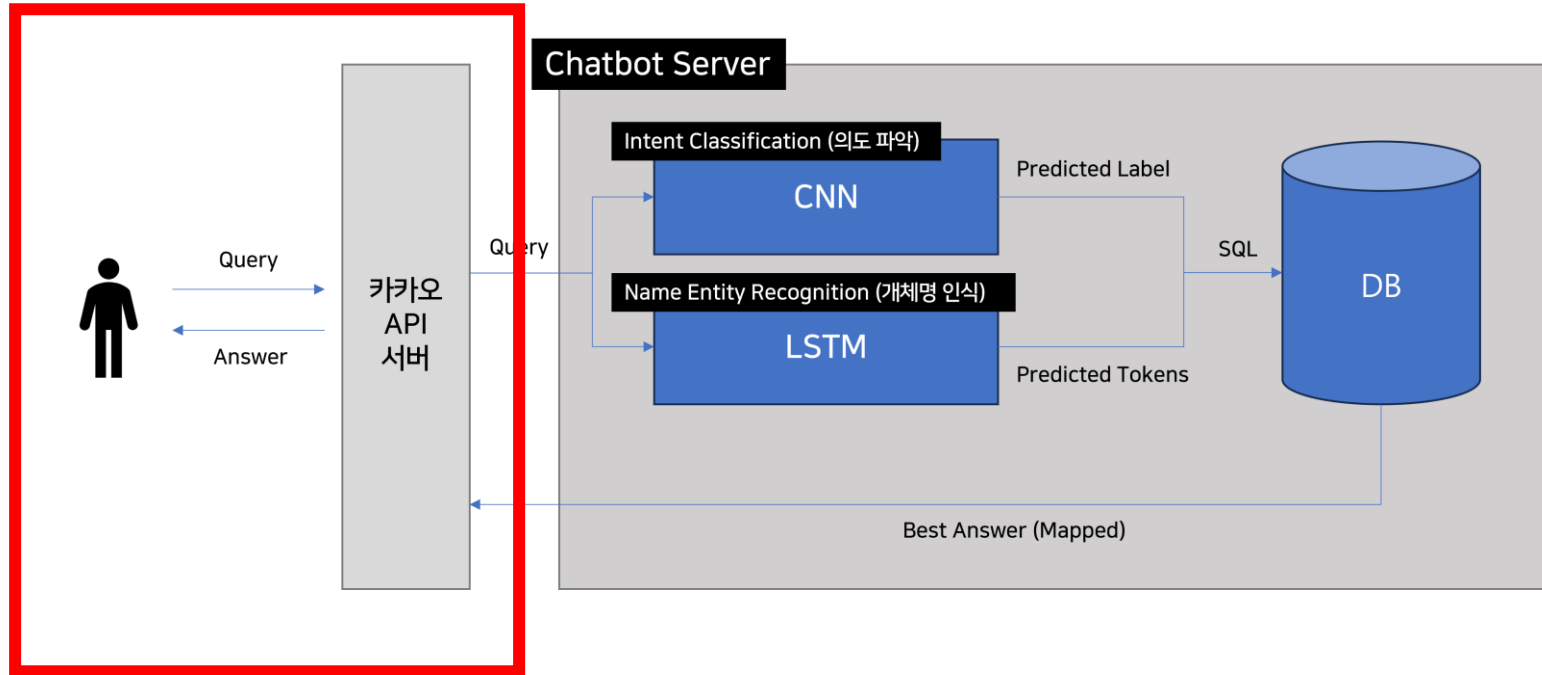
Python Flask 웹 프레임워크를 이용한 Server 개발

3-1. 개발 진행예정



Client - Server 테스트

3-1. 개발 진행예정



카카오 API 서버, Chatbot Server 연결, 테스트 서비스 배포

4. 한계점 및 개선사항

한계점

1. 데이터 부족, Labeled dataset do not exist.
2. 모델의 정확성 (Mapping 방법으로 인한 정확성 저하 문제)

4. 한계점 및 개선사항

개선사항

1. 지속적인 데이터 구축, 데이터 Annotation 진행
2. 최신 모델 (BERT, GPT) 등을 이용한 모델 개선